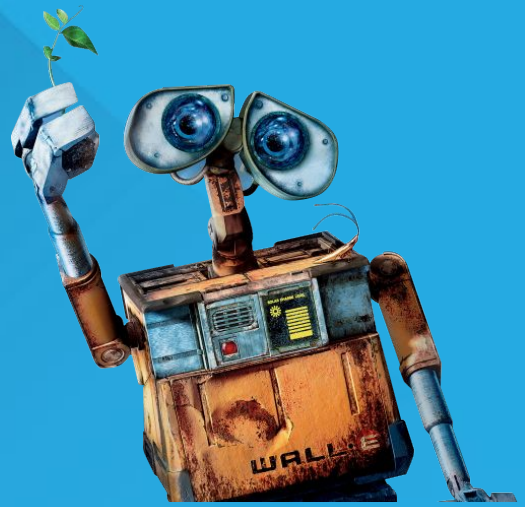
Sentiment Analysis: Twitter Hashtag Disinformation Analyzer



Murtadha Marzouq

Survey Research

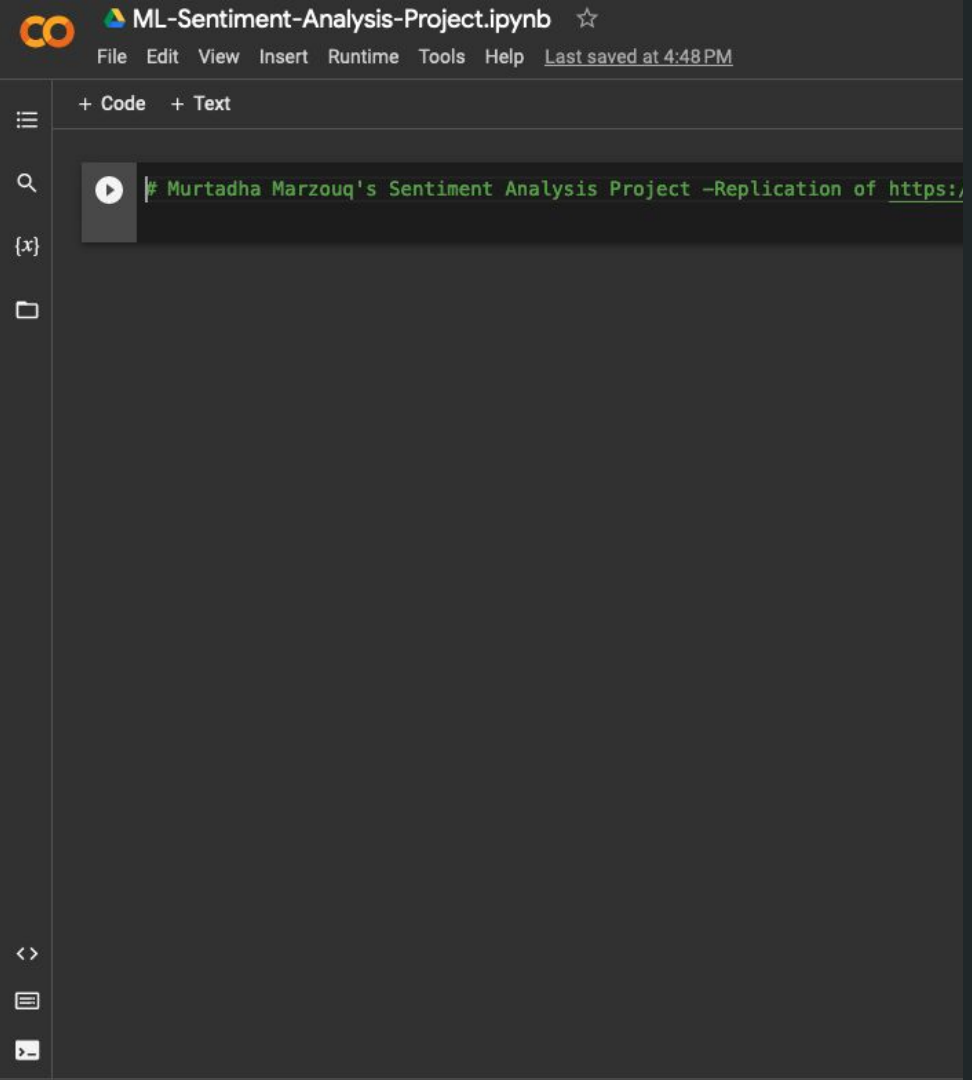
Problem: Using Sentiment Analysis to help identify Disinformation to avoid untrustworthy sources.

Goal: I will be creating an interactive notebook where it ranks Hashtags by safety scale (0-10) using NLP and Sentiment Analysis.

Timeline:

A. Research & Theory (1 week)
B. Practical Analysis (1 week)
C. Demo & Reproduction of existing work (1 Week)

Github Repo: https://github.com/MurtadhaM/ML-5156



Summary & Overview:

Previous Work:

Lanier, Heather D., et al. "Analyzing COVID-19 Disinformation on Twitter Using the Hashtags #scamdemic and #plandemic: Retrospective Study." PloS One, vol. 17, no. 6, 2022, pp. e0268409–e0268409, https://doi.org/10.1371/journal.pone.0268409

Reproduction Steps:

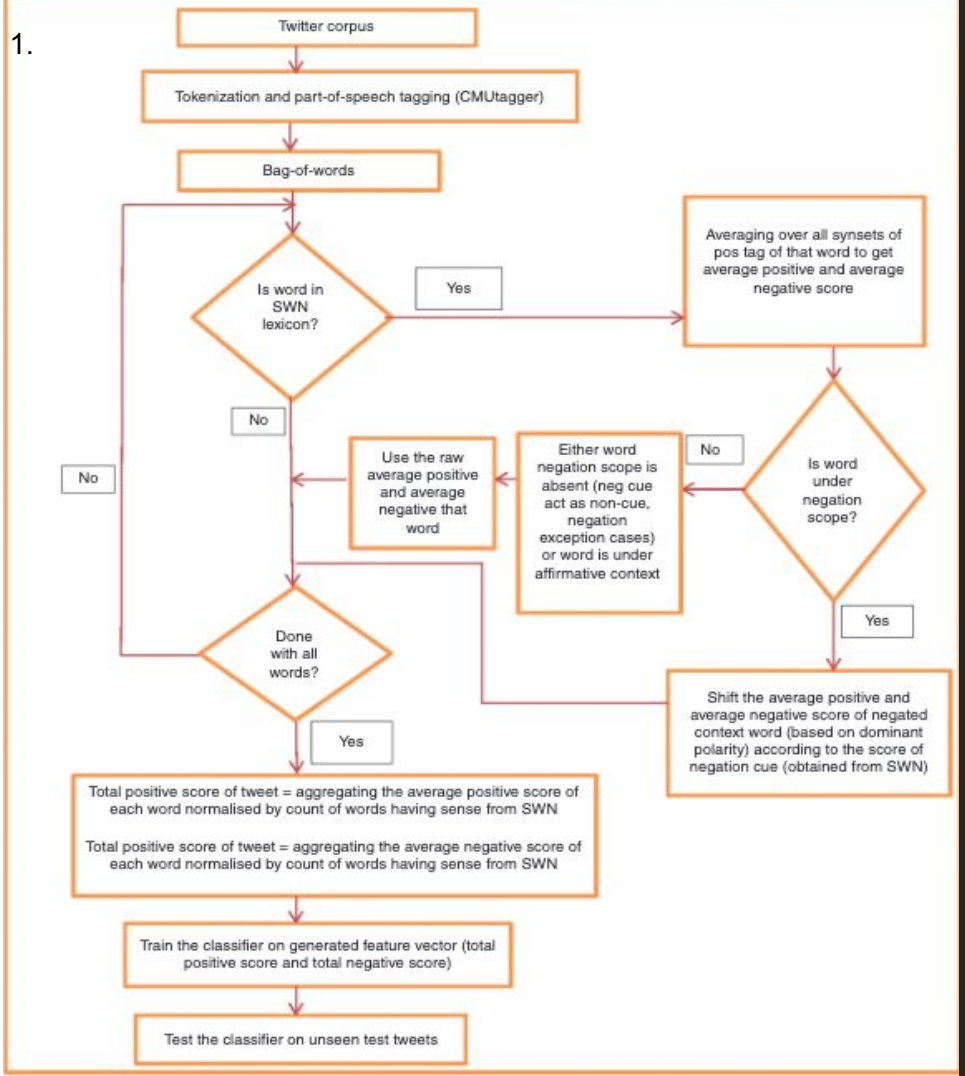
- Extracting Data from Twitter
- Loading & Clean up the Data
- Using LDA Model to perform Risk Analysis
- Present the Data

Methods Used:

- TWINT too to extract the data
- Pandas to manage the data
- LDA from Gen Sim library in Python
- Matplotlib & Seaborn to create graphs.

Notebook: https://colab.research.google.com/drive/10BrztOdvtKxCPL_WbCMT7gDABQVOOefs?usp

1.



Research Summary

Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic

1. Collect a dataset consisting of lung CT scans from patients with and without lung cancer.
2. Preprocess the data to remove noise and enhance image features.
3. Extract image features using a technique called local binary pattern (LBP).
4. Train machine learning models such as decision tree, & K-nearest neighbors.
5. Evaluate the performance of the trained models using metrics such as accuracy,
6. Interpret the results and assess the clinical relevance of the proposed method

Gupta, Itisha, and Nisheeth Joshi. "Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic." Journal of Intelligent Systems, vol. 29, no. 1, 2020, pp. 1611–25, https://doi.org/10.1515/jisys-2019-0106.

Research Summary

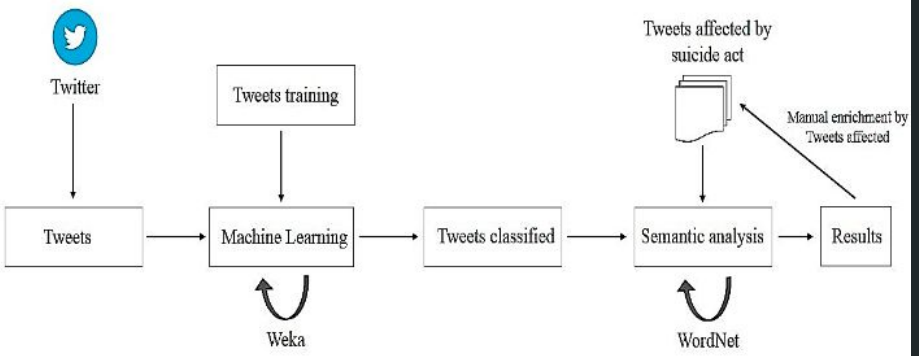


Fig. 1. Architecture of our methodology work of suicide detection.

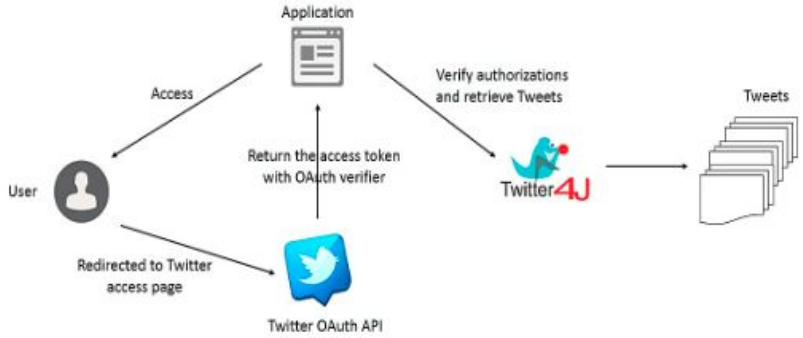


Fig. 2. Architecture for Tweet collection from Twitter using Twitter4J.

Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks

1. Collect data from Twitter using Twitter4J API.
2. Develop a vocabulary associated with suicide to analyze tweets related to suicide.
3. Use machine learning algorithms such as Support Vector Machines (SVM), Maximum Entropy and Naive Bayes for classification.
4. Use semantic analysis based on WordNet to determine the polarity of sense to users.
5. Use Weka tool for sentiment analysis based on machine learning algorithms to extract useful information from Twitter data.
6. Compute semantic analysis between tweets in training set and tweets in data set based on WordNet.
7. Verify the effectiveness of performance in terms of accuracy and precision on semantic sentiment analysis related to suicidal ideation.

Birjali, Marouane, et al. "Machine Learning and Semantic Sentiment Analysis Based Algorithms for Suicide Sentiment Prediction in Social Networks." Procedia Computer Science, vol. 113, 2017, pp. 65–72, https://doi.org/10.1016/j.procs.2017.08.290.

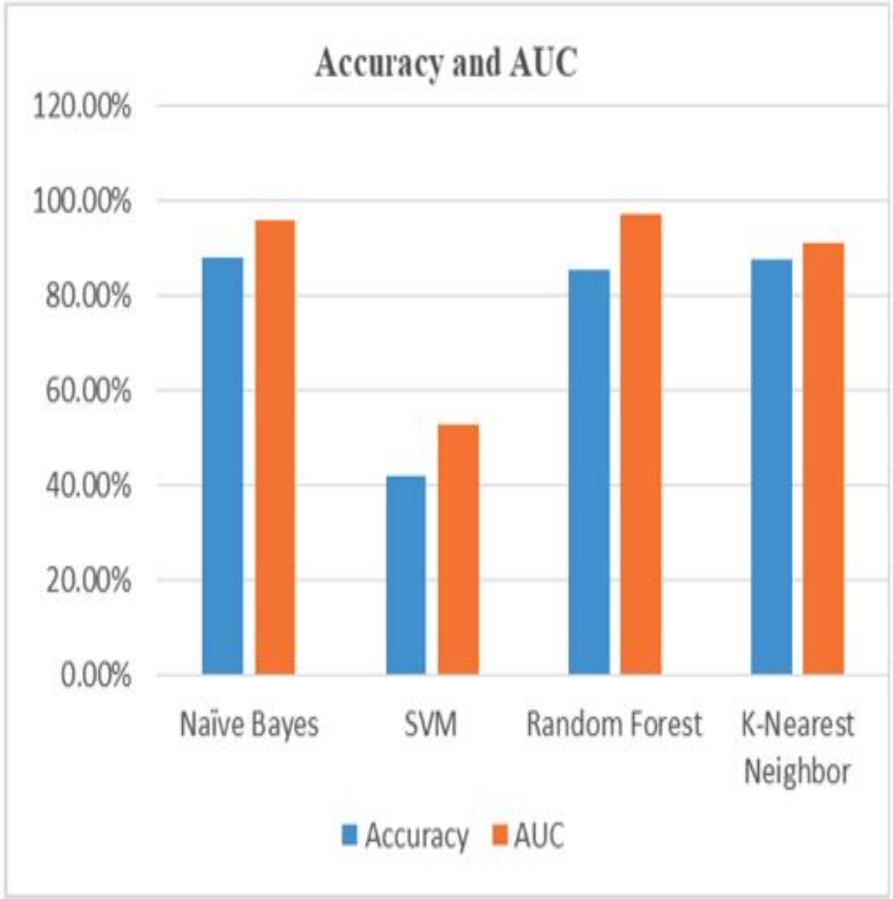


Figure 9. Graphical results analysis.

Research Summary

Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning

1. Filter and analyze Twitter data using natural language processing techniques to detect sentiment polarity based on emotional words.
2. Normalize the dataset to be used by machine learning algorithms.
3. Use natural language processing techniques like word tokenization, stemming and lemmatization, POS tagger, NER, and parser to extract emotions from each tweet.
4. Implement the approach using Python programming language and NLTK.
5. Use a derived algorithm to extract emotional words using WordNet with its POS for the word in a sentence, and assign sentiment polarity using SentiWordNet or a lexicon-based method.
6. Analyze the resultant polarity using naïve Bayes, SVM, KNN, and random forest machine learning algorithms.
7. Visualize the results on the Weka platform.
8. Naïve Bayes gives the best accuracy of 88.17% while random forest gives the best AUC of 0.97.

Patel, Ravikumar, and Kalpdrum Passi. "Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning." IoT, vol. 1, no. 2, 2020, pp. 218–39, https://doi.org/10.3390/iot1020014.